The eyes follow the sound: Measuring speech perception with eye-tracking

Sónia Frota[1], Cátia Severino[1], Joseph Butler[1], Cláudia B. Lima[1,2] & Marina Vigário[1]

[1]University of Lisbon, [2]Santa Maria Hospital - CHLN

17th European Conference on Eye Movements, 11-16 August 2013, Lund, Sweden

Introduction

- Infant speech perception is known to undergo critical changes before the first year of life: young infants are able to discriminate non-native speech contrasts, whereas older infants and adults lose this ability by attuning to the native language (Werker & Tees 1984/2002; Kuhl et al. 2005; Saffran et al. 2006; Tsao et al. 2006). This dramatic change in non-native speech perception has been shown to predict later language development (Kuhl et al. 2005; Rivera-Gaxiola et al. 2005).
- Much of this research was conducted using the Conditioned Head Turn (CHT) procedure (Werker et al. 1997), which is a visually reinforced infant speech discrimination paradigm that assesses the discrimination of sound categories on the basis of the presentation of multiple exemplars. For adults, AB or ABX discrimination tasks have been used (e.g., Tsao et al. 2006).
- Unlike the head-turning dependent measure, eye-movements are very fast, have low metabolic cost, and can be automatically detected by eye-tracking systems. Furthermore, eye movements can be used as a dependent measure in both infant and adult studies. Anticipatory eye movements have been shown to cue the linguistic processing of sentences by adults (Altmann & Kamide 2007; Kukona et al. 2011), and to measure the categorization of visual and auditory stimuli, as well as the learning of audio-visual contingencies by infants as young as 6 months (McMurray & Aslin 2004; Shulka et al. 2011; Bjerva et al. 2011).

The present study

Goal: To develop an eye-tracking version of CHT by exploring ***anticipatory eye movement*** (AEM) triggered by auditory stimuli.

- On the basis of an audio-visual contigency and conditioned visual reinforcement, we used AEM to assess the discrimination of non-native speech sound categories by European Portuguese (EP) adults and infants.
- The Hindi sound contrast /d'a/ vs. /da/, studied by Werker et al. for English, was used.
- Three experiments were conducted:
  - Experiment 1 examined adult perception in an ABX task;
  - Experiment 2 examined adult perception in an AEM paradigm;
  - Experiment 3 examined infant perception in an AEM paradigm.
- Eye-movement was automatically detected by an SMI RED250 system.

Experiment 1: ABX task

Method

Participants

18 native Hindi speakers (8 females)
18 native EP speakers (12 females)

Stimuli

3 natural speech pairs of tokens of the Portuguese native contrast /pa/ - /ka/, produced by a female speaker, were used in the training phase.
6 natural speech pairs of tokens of the Hindi contrast /d'a/ - /da/ were used in the test phase. The Hindi stimuli were produced by a male native Hindi speaker.

Procedure

3 pairs of tokens of the Portuguese contrast /pa/ - /ka/ were used in the training phase, which consisted of 12 trials (with response feedback).
6 pairs of tokens of the Hindi contrast /d'a/ - /da/ were used in the test phase, which consisted of 7 blocks with 24 trials each (no feedback). Participants had to respond within 2000ms after hearing the X sound.

Results and discussion

Portuguese-speaking adults scored lower than Hindi speaking adults on Hindi /d'a/ - /da/ discrimination, as shown by their larger error rate (35.1% vs. 20.3%). A one-way ANOVA revealed a significant difference between the two groups ($F(1,35) = 10.47$, $p < .01$, $\eta^2 = .24$). The results of Exp.1 indicate that language experience affects adult discrimination of non-native contrasts.

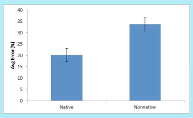

Figure 1: Error rate across the two language groups

Experiments 2 and 3: AEM paradigm

Exp. 2: Method

Participants

19 native EP speakers (M=29; range 21-38; 12 female)

Stimuli

The speech sounds used in Exp.1: The native contrast /pa/ - /ka/ (labial vs. velar plosive) in the training phase; The Hindi contrast /d'a/ - /da/ (retroflex vs. dental place of articulation) in the test phase.

Exp. 3: Method

Participants

10 EP infants (M=8 mos, 26 days; range 5 m 2 days – 20 m 23 days; 7 female)

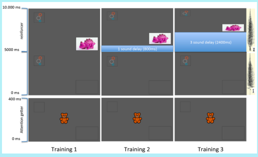

Figure 2: AEM paradigm: Training phase trial setup

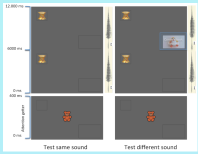

Figure 3: AEM paradigm: Test phase trial setup

Procedure

Training phase: 3 blocks of 4 trials each. Each trial began with an attention getter at the center of the screen (trigger AOI, 400 ms), followed by a dynamic visual stimulus at the top left side coincident with the presentation of the speech sound file. Each sound file consisted of 6 tokens of /pa/ followed by 6 tokens of /ka/, with an ISI of 1500 ms. In block 1, the onset of the 1st /ka/ sound is aligned with an attractive video reinforcer (VAOI) at the bottom right side of the screen. In blocks 2 and 3, VAOI onset was delayed relative to the onset of the sound change (Fig. 2). Two different reinforcers were used.

Test phase: 24 trials (12 no change trials: /d'a/ only or /da/ only; 12 change trials: /d'a/ > /da/ or /da/ > /d'a/). In the change trials, VAOI appears after 400 ms looking time to the correct side while sound 2 is playing (Fig. 3). Two reinforcers different from those in the training were used in the test phase.

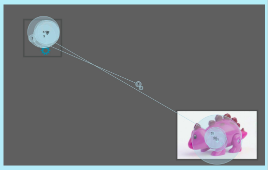

Figure 4: Eye-gaze during a training trial, as shown by the main SMI Begaze display

Experiments 2 and 3: AEM paradigm

Exp. 2: Results and discussion

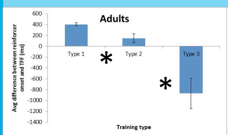

Figure 5. Average difference (ms) between reinforcer onset (VAOI) and TTF within VAOI, across the 3 training blocks

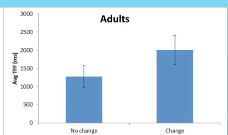

Figure 6. Average TTF (ms) within VAOI in the test trials (fixations were found in only 22 no change and 24 change trials out of 114 trials)

Time to first fixation (TTF) within the VAOI was used as the dependent measure. AEM is obtained if TTF occurred in the time window between the onset of the sound change and the onset of the video reinforcer (VAOI) in training blocks 2 and 3.

- A repeated measures ANOVA revealed a significant effect of Training block ($F(2,26) = 17.89$, $p < .001$, $\eta 2 = .58$)

- Non-parametric Wilcoxon signed ranks revealed significant differences between Training blocks 1 and 2 ($z = 2.46$, $p < .05$), blocks 1 and 3 ($z = 3.17$, $p < .01$) and blocks 2 and 3 ($z = 3.3$, $p < .01$).

-The results of the ***training phase*** indicate that subjects exhibit AEM during the last training block (7 out of 19 already show it in block 2).

- However, in the ***test phase*** subjects showed no AEM to the non-native sound contrast (A repeated measures ANOVA revealed no effect on Trial Type ($F(1,112) = 2.17$, $p = .14$, $\eta^2 = .02$), no effect of order/type of stimulus ($F(1,112) < 1$) and no interaction ($F(1,112) < 1$). This indicates they were not able to discriminate the non-native contrast, replicating the results of the ABX experiment (Exp. 1).

Exp. 3: Results and discussion

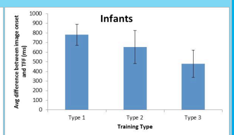

Figure 7. Average difference (ms) between reinforcer onset (VAOI) and TTF within VAOI, across the 3 training blocks

- Infants showed no AEM in the training (A repeated measures ANOVA revealed no effect of Training block ($F(2,18) = 1.43$, $p = .27$, $\eta^2 = .14$; Non-parametric Wilcoxon signed ranks revealed no significant differences; Only 1 out of 10 show AEM in block 2 and in block 3).

- However, an effect in the right direction emerged: Infants are looking quicker to VAOI in block 2 than in block 1, and in block 3 than in block 2. Possibly, the amount of training was not enough to trigger antecipatory looking (18 to 30 trials in previous studies)

Discussion

- We developed an eye-tracking version of CHT by exploring ***anticipatory eye movement*** triggered by auditory stimuli to assess speech discrimination by adults and infants.
- Adults were found to present AEM after 12 training trials, and failed to discriminate a non-native speech sound contrast in the test phase. This result replicates findings from an ABX discrimination task, supporting the effect of language experience on adult discrimination of non-native contrasts.
- Our results from Exp.2 thus suggest that the AEM paradigm developed is successful to assess discrimination of speech sound categories.
- Infants did not present AEM, although there is evidence that they were learning to predict image appearance from the sound heard. Different factors may have affected infants' behavior: (i) amount of training (12 training trials were not enough); (ii) attractiveness of the visual stimulus that holds infants' attention during sound presentation; (iii) diversity of the visual reinforcer (only two different videos were used).
- The impact of these factors will be examined in future research.

References

Altmann, G.T.M & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57, 502-518.

Bjerva, J., Marklund, E, Engdahl, J. & Lacerda, F. (2011). Anticipatory Looking in Infants and Adults. *Proceedings of EyeTrackBehavior*, Stockholm University.

Kukona, A., Fang, S., Aicher, K.A., Chen, H. & Magnuson, J.S. (2011). The time course of anticipatory constraint integration. *Cognition*, 119, 23-42.

Kuhl, P., Conboy, B.T., Padden, D., Nelson, T. & Pruitt, J. (2005). Early Speech Perception and Later Language Development: Implications for the "Critical Period". *Language Learning and Development*, 1(3&4), 237-264.

McMurry, B. & Aslin, R.N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, 6(2), 203-229

Rivera-Gaxiola, M., Silva-Pereyra, J. & Kuhl, P.K. (2005). Brain potentials to native- and non-native speech contrasts in seven and eleven-month-old American infants. *Developmental Science*, 8, 167-172

Saffran, J.R., Werker, J.F. & Werner, L.A. (2006). The infant's auditory world: Hearing, speech and the beginnings of language. In R. Siegler & D. Kuhn (Eds.), *Handbook of child development* (Vol. 6, pp. 58-108). New York: Wiley.

Shulka, M., Wen, J., White, K.S. & Aslin, R.N (2011). SMART-T: A system for novel fully automated anticipatory eye-tracking paradigms. *Behavior Research Methods*, 43(2), 384-98.

Tsao, F., Liu, H. & Kuhl, P.K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *Journal of the Acoustical Society of America, 120 (4),* 2285-2294.

Werker, J.F. & Tees, R.C. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 25,* 121-133.

Werker, J., Polka, L. & Pegg, J (1997). The Conditioned Head Turn Procedure as a Method for Testing Infant Speech Perception. *Early Development and Parenting, 6,* 171-178.

EXCL/MHC-LIN/0688/2012 – EBELa Project, FCT

Contact email: sonia.frota@mail.telepac.pt

Acknowledgments:
We thank the Embassy of India and Shiv Kumar Singh for recruting the Hindi subjects for Experiment 1. Shiv has kindly recorded the Hindi stimuli.

IF INVESTIGADOR FCT
FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

Lisbon Baby Lab
Universidade de Lisboa